



UNIVERSITY OF  
**GEORGIA**

# Tuning indirect predictions based on SNP effects from ssGBLUP

**Daniela Lourenco**

A. Legarra, S. Tsuruta, D. Moser, S. Miller, I. Misztal

# Why Indirect predictions?

- Interim evaluations
  - Between official runs
- Not all genotyped animals are in the evaluations
  - Animals with incomplete pedigree increase bias and lower  $R^2$
- Commercial products
  - e.g. GeneMax for non-registered animals

# Indirect predictions in ssGBLUP

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{W} + \mathbf{H}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix}$$

$$\mathbf{DGV} = \mathbf{Z}\hat{\mathbf{a}}$$

SNP effects

GEBVs

$$\hat{\mathbf{a}} = \lambda \mathbf{D} \mathbf{Z}' \mathbf{G}^{-1} \hat{\mathbf{u}}$$

$$\mathbf{GEBV}_{\text{young}} = w_1 \mathbf{PA} + w_2 \mathbf{DGV} - w_3 \mathbf{PP}$$

$$\mathbf{GEBV}_{\text{young}} \approx \mathbf{DGV} = \mathbf{Z}\hat{\mathbf{a}}$$

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

$$\mathbf{G}_{\text{APY}}^{-1}$$

Lourenco et al., 2015

# Problems with Indirect predictions

## Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus<sup>1</sup>

D. A. L. Lourenco,<sup>\*2</sup> S. Tsuruta,<sup>\*</sup> B. O. Fragomeni,<sup>\*</sup> Y. Masuda,<sup>\*</sup> I. Aguilar,<sup>†</sup>  
A. Legarra,<sup>‡</sup> J. K. Bertrand,<sup>\*</sup> T. S. Amen,<sup>§</sup> L. Wang,<sup>§</sup> D. W. Moser,<sup>§</sup> and I. Misztal<sup>\*</sup>

© 2015 American Society of Animal Science. All rights reserved. J. Anim. Sci. 2015.93:2653–2662  
doi:10.2527/jas2014-8836

$$\text{COR}(\widehat{\text{GEBV}}, \mathbf{Z}\hat{\mathbf{a}}) > 0.99$$

$$\text{Avg}(\widehat{\text{GEBV}}) \approx 100 \quad \neq \quad \text{Avg}(\mathbf{Z}\hat{\mathbf{a}}) \approx 0$$

# Objectives

- 1) Fine-tune indirect predictions to be compatible with GEBV
  
- 2) Investigate whether SNP effects are accurate when APY is used
  - Possibly use subset of core animals

# Dataset

- American Angus Association
  - 8.2M animals in pedigree
  - 6.2M birth weight (BW)
  - 6.8M weaning weight (WW)
  - 3.4M post-weaning gain (PWG)
  - 81k genotyped
    - born 1977-2012: 66k
    - born 2013-2014: 15k
- Complete
  - Phenotypes up to 2012
  - Genotypes up to 2014 (81k)
- Reduced
  - Phenotypes up to 2012
  - Genotypes up to 2012 (66k)
- 3-trait with mat and mpe
  - Results for PWG

# Accuracy of SNP effects from $\mathbf{G}_{APY}^{-1}$ or $\mathbf{G}_{CC}^{-1}$

$$\hat{\mathbf{a}}_{\mathbf{G}} = \lambda \mathbf{D} \mathbf{Z}' \mathbf{G}^{-1} \hat{\mathbf{u}}$$

$$\hat{\mathbf{a}}_{\mathbf{G}_{APY}^{-1}H} = \lambda \mathbf{D} \mathbf{Z}' \mathbf{G}_{APY\_high\_reliability}^{-1} \hat{\mathbf{u}}_{APY}$$

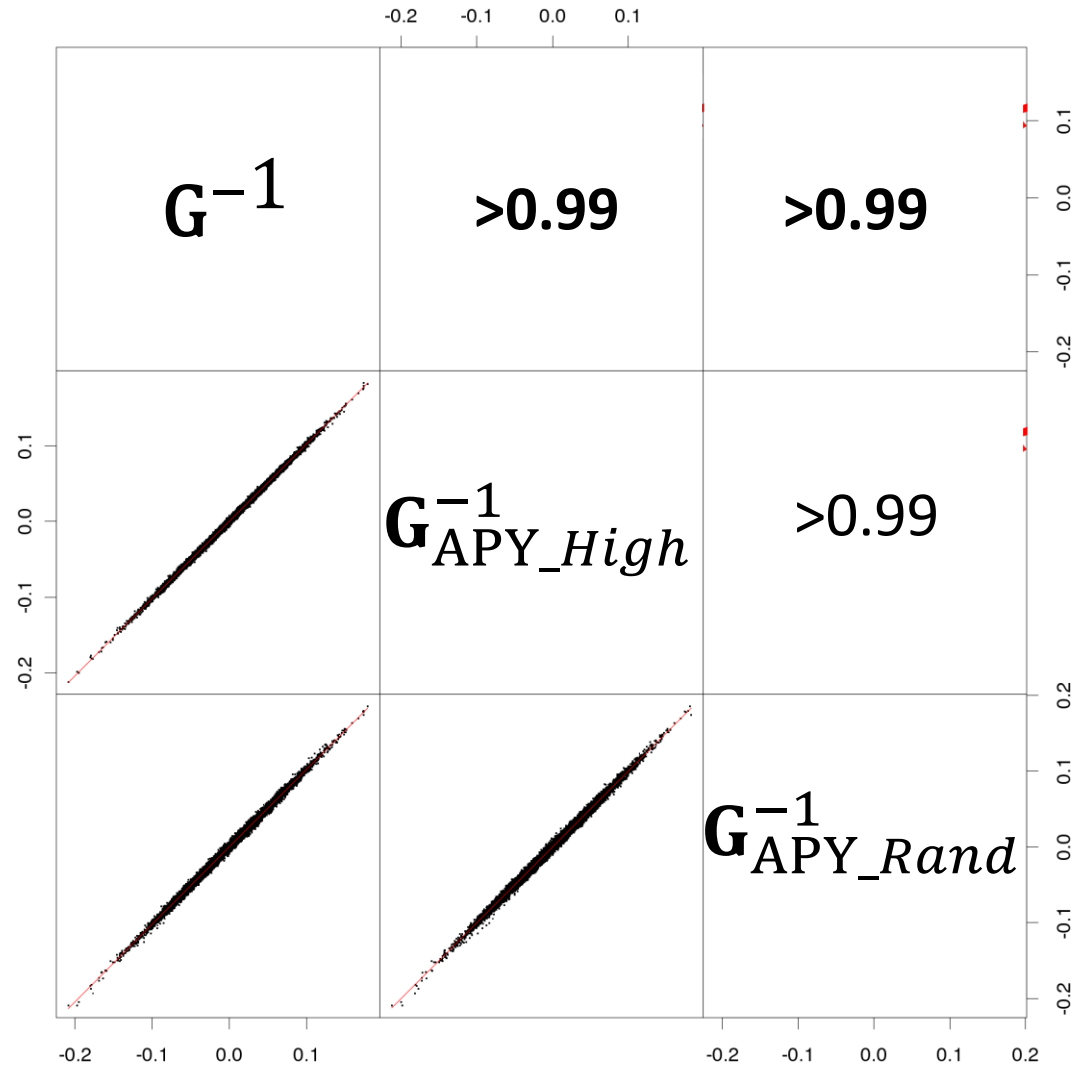
$$\hat{\mathbf{a}}_{\mathbf{G}_{APY}^{-1}R} = \lambda \mathbf{D} \mathbf{Z}' \mathbf{G}_{APY\_random}^{-1} \hat{\mathbf{u}}_{APY}$$

$$\hat{\mathbf{a}}_{\mathbf{G}_{CC}^{-1}H} = \lambda \mathbf{D} \mathbf{Z}' \mathbf{G}_{CC\_high\_reliability}^{-1} \hat{\mathbf{u}}_{APY}$$

$$\hat{\mathbf{a}}_{\mathbf{G}_{CC}^{-1}R} = \lambda \mathbf{D} \mathbf{Z}' \mathbf{G}_{CC\_random}^{-1} \hat{\mathbf{u}}_{APY}$$

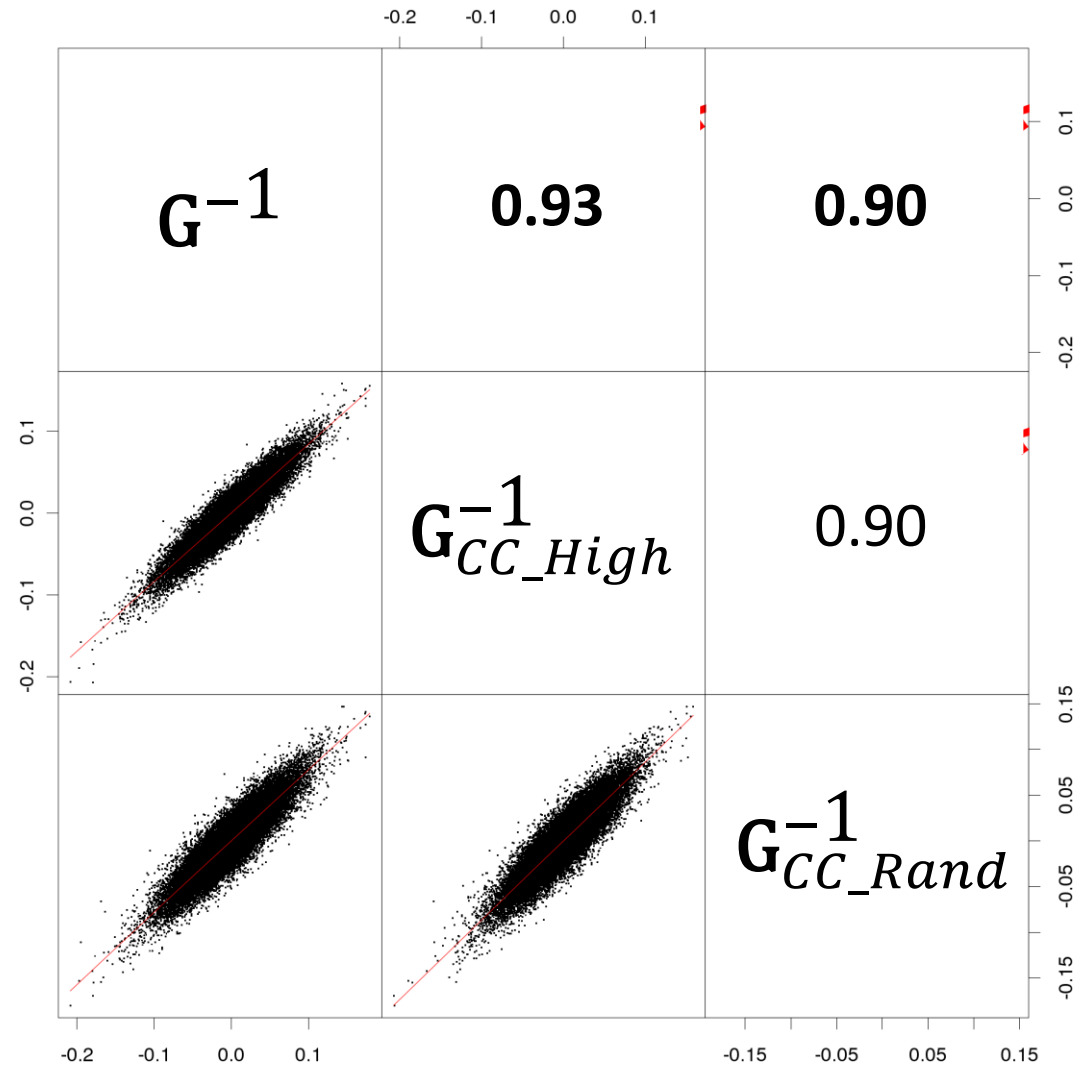
- Correlation between SNP effects
- Correlation between  $\mathbf{Z}\hat{\mathbf{a}}$

# Statistics for SNP effects

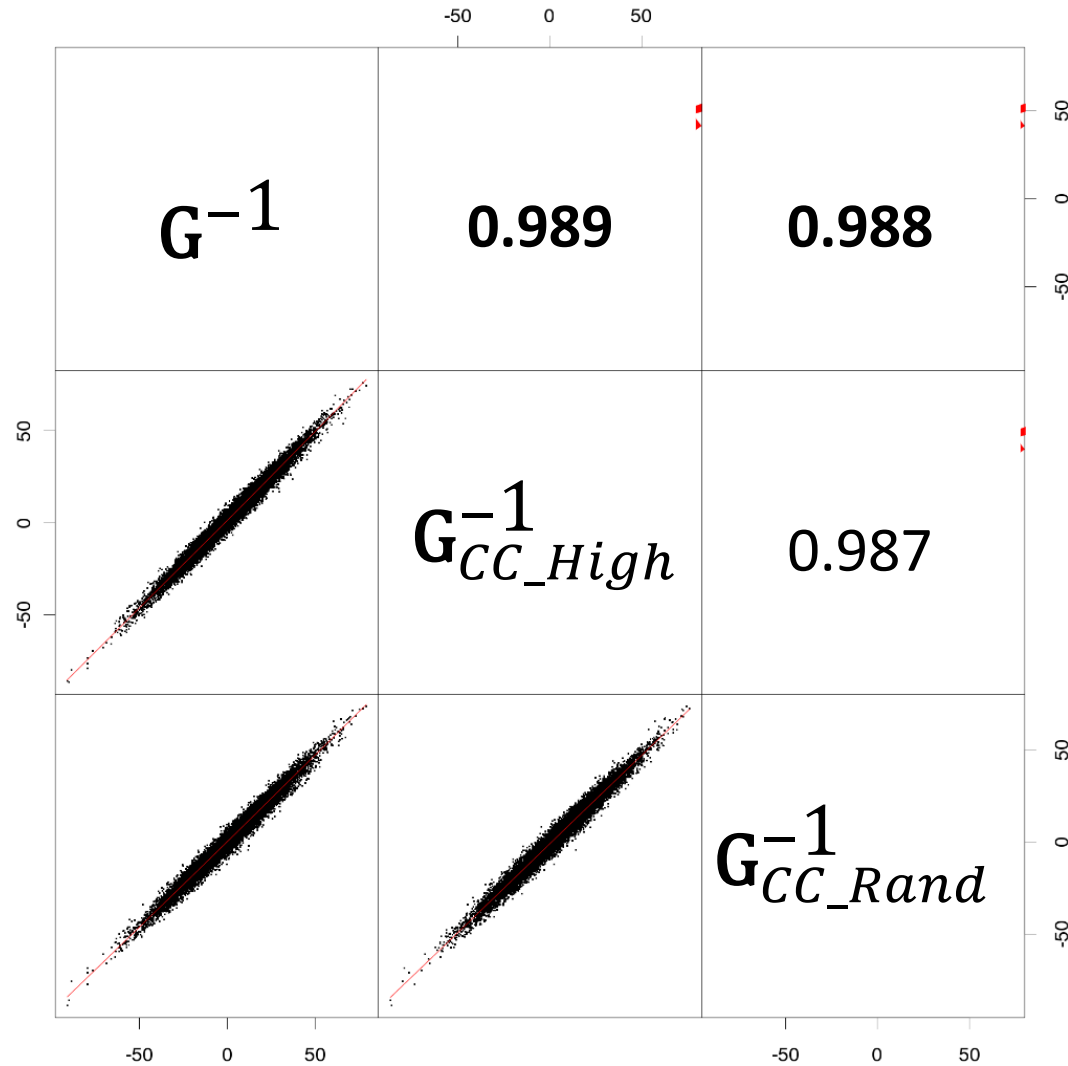




# Statistics for SNP effects



# Statistics for $\mathbf{Z}\hat{\mathbf{a}}$



# Fine-tuning indirect predictions from ssGBLUP

Understanding genetic and genomic bases

- Base of BLUP: *founders of the pedigree*
- Base of GBLUP: *genotyped* animals
- Base of SSGBLUP: Vitezica et al. (2011) modeled as a mean in genotyped animals
  - $p(\mathbf{u}_g) = N(\mathbf{1}\mu, \mathbf{G})$
  - $\mu = (\text{Pedigree base}) - (\text{Genomic base})$

# Fine-tuning indirect predictions from ssGBLUP

1) Formula in Legarra (2017)

$$\hat{\mathbf{u}}_{ip} = \hat{\mu} + 0.95\mathbf{Z}\hat{\mathbf{a}} + 0.05 \hat{\mathbf{u}}_{parents}$$

SNP and  
pedigree  
fractions

2) Double fitting

a) fit a regression using genotyped animals in the evaluation

$$DGV_{eval} = b_0 + b_1\mathbf{Z}\hat{\mathbf{a}}$$

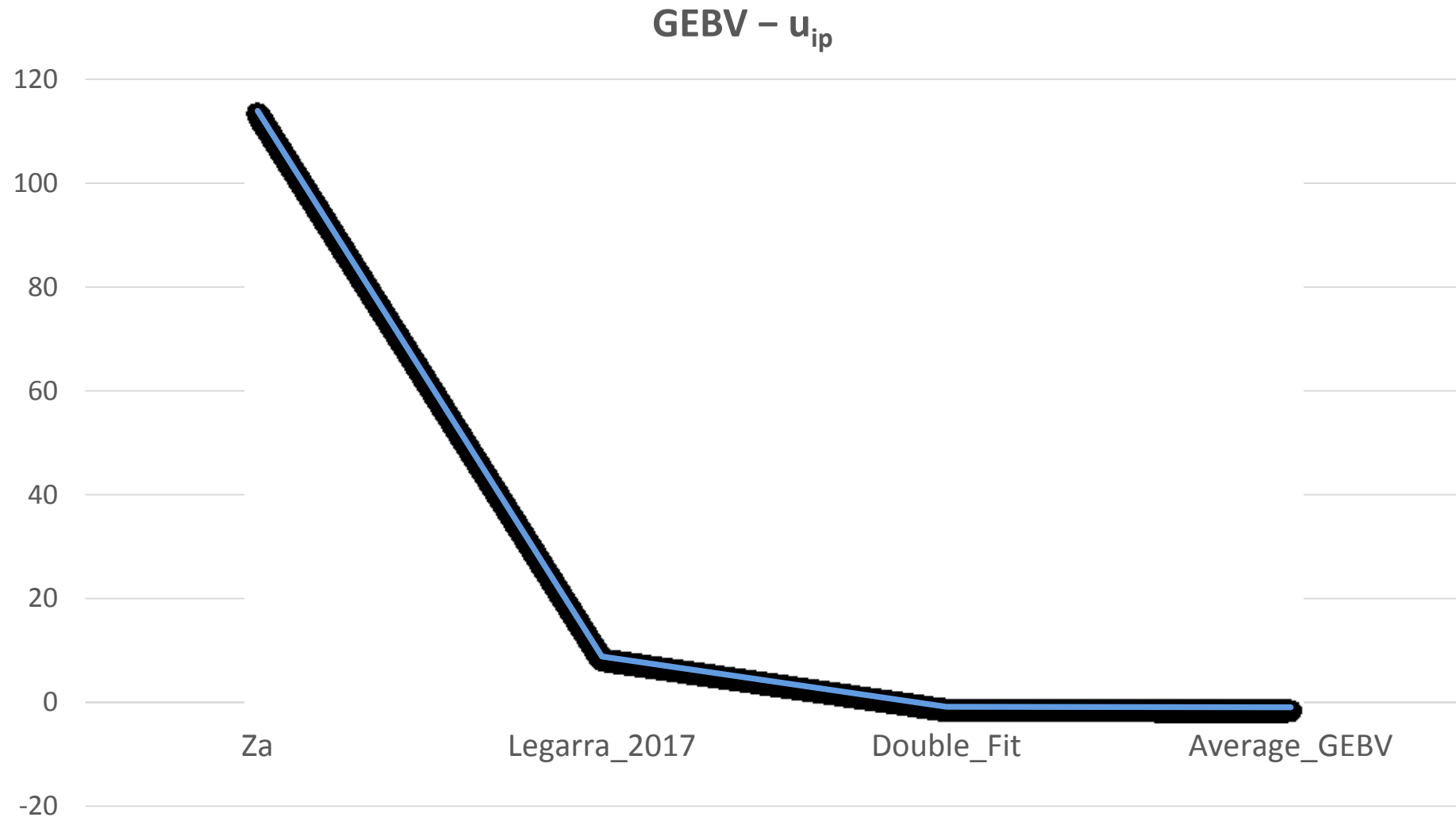
b) apply regression for indirectly predicted animals

$$\hat{\mathbf{u}}_{ip} = b_0 + b_1\mathbf{Z}\hat{\mathbf{a}}$$

3) Add average GEBV

$$\hat{\mathbf{u}}_{ip} = \overline{GEBV}_{eval} + \mathbf{Z}\hat{\mathbf{a}}$$

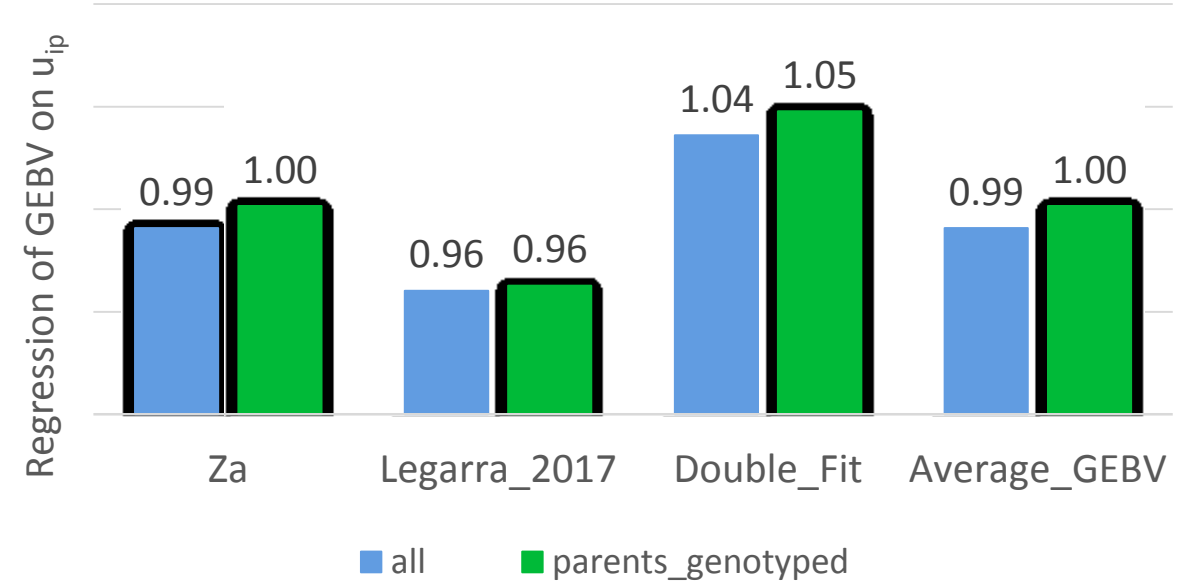
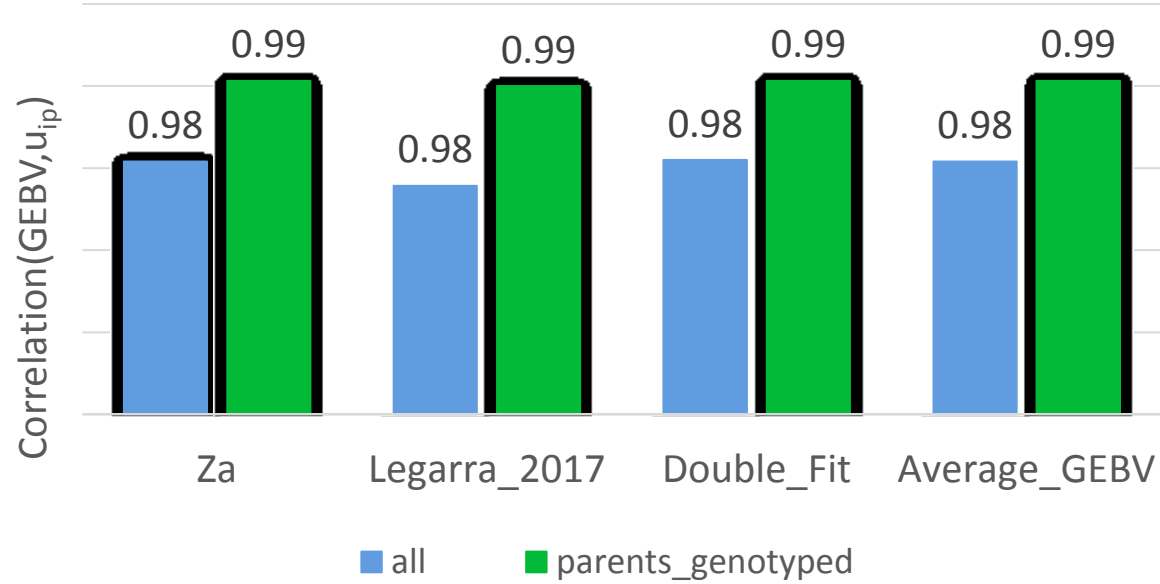
# Bias of indirect predictions



# Correlation

&

# Regression Coefficient



# Fine-tuning indirect predictions in ssGBLUP

$$E(\hat{\mathbf{u}}|\hat{\mathbf{a}}) = \mu + \mathbf{Z} \frac{\mathbf{1}}{2 \sum p(1-p)} \left( \mathbf{I} \frac{\mathbf{1}}{2 \sum p(1-p)} \right)^{-1} (\hat{\mathbf{a}} - 0)$$

$$E(\hat{\mathbf{u}}|\hat{\mathbf{a}}) = \mu + \mathbf{Z}\hat{\mathbf{a}}$$

$\approx$

$$E(\hat{\mathbf{u}}|\hat{\mathbf{a}}) = \overline{GEBV} + \mathbf{Z}\hat{\mathbf{a}}$$

# Final Remarks

- Indirect predictions are unbiased after corrections
  - Can be used as interim evaluation
- Indirect predictions based on core animals are slightly less accurate
  - Reduction in computing time (no  $\mathbf{G}_{nC}^{-1}$  and  $\mathbf{G}_{n\bar{n}}^{-1}$ )
- SNP effects from ssGBLUP may be useful for SNP MACE



# Acknowledgements

